

# Brain Language: Uncovering Functional Connectivity Codes

Victor M. Vergara  
Medical Image Analysis  
Laboratory  
The Mind Research Network  
Albuquerque, NM, USA  
[vvergara@mrn.org](mailto:vvergara@mrn.org)

Vince D. Calhoun  
Medical Image Analysis  
Laboratory  
The Mind Research Network  
Albuquerque, NM, USA  
[vvergara@mrn.org](mailto:vvergara@mrn.org)

**Abstract**— The functional connectivity within a specific set of brain networks (or domain) can assume different configurations known as domain states that change with time. Recently, we proposed an information theoretical framework that models the finite set of domain states as elements of an alphabet. Significant bits of information have been found to be shared among domains, but specific domain codification was not explored. This work describes a method to identify code words used to transmit and receive information between the cerebrum and the cerebellum based on dynamic domain connectivity estimated from functional magnetic resonance imaging (fMRI). Following the theory of jointly typical sets, the developed method identifies the codeword length and the specific combination of domain states on each codeword. Resting state fMRI data was taken from 121 subjects with no significant age difference between males and females. Group independent component analysis was utilized to identify important brain networks and group them in a cerebellum and six other domains representing the cerebrum. The amount of information between the cerebellum, the executive control and sensorimotor domains showed a statistically significant number of bits. The proposed method quantified specific temporal sequences of domain states encoded within bits shared between cerebellum and cerebrum.

**Keywords**—dynamic functional connectivity, mutual information, information theory, brain, cerebellum

## I. INTRODUCTION

Technological advances in the analysis of functional magnetic resonance imaging (fMRI) data have provided knowledge about the way brain areas are connected to each other [1]. One important discovery suggests that our brains have evolved into a connected network consisting of sparse local clusters with few long-range connections among clusters [2]. These clusters can be found in resting state fMRI defining brain domains (sets of related brain areas) grouped according to functional associations [1, 3, 4]. Two examples of these domains are the set of brain areas in charge of sensorial-motor (SEN) functions and the collection of functionally different sections of the cerebellum (CER). An informational theoretical framework has been proposed recently to study the communication between domains by estimating the number of bits required for this communication [5]. Although the existence of bits of information shared among domains can be estimated, further development is required to determine what is the underlying code used for this communication.

A common way of studying the resting state brain, fMRI scans obtained during rest without a particular task, is to analyze coactivations among brain areas. Different but replicable patterns of coactivations have been found to define the brain state at a given moment in time [6]. Several techniques have been developed to identify and track these dynamic state changes [7, 8]. Instead of whole brain analysis, this work is concerned with the dynamic states of individual domains estimated by utilizing the dynamic functional domain connectivity (dFDC) previously proposed [9]. Fig. 1 synthesizes the dFDC procedure for domains CER and SEN. Temporal activation information extracted from the components of two domains of interest is used to estimate short lived correlations within a temporal window at different points in time. This procedure result in a sequence of correlation matrices defined at each time point. The connectivity patterns from the correlation matrix forms a finite set that shuffles sequentially in time. The dFDC analysis identifies these patterns and determines their temporal sequence. The probability distribution of estimated dFDC states is then used to calculate information theory measures that characterize the domain connectivity [5].

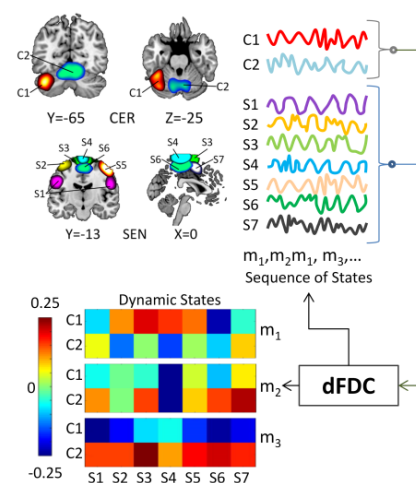


Fig. 1. Summary of the dFDC procedure. A pair of domains are selected. Then a sliding window correlation and clustering are performed to discover dFDC states and their temporal sequence of appearing.

This work utilizes outcomes from the dFDC analysis to decode sequences of states that determine the communication among domains. We have previously proposed a method to estimate entropy and mutual information from dFDC [5]. Borrowing from communication theory, the existence of mutual information indicates an underlying messaging system with its own coding structure. In the following, we make a first attempt to uncover this coding by selecting the cerebrum and cerebellum as two main structures in the central nervous system.

## II. CODES AND JOINTLY TYPICAL SETS

Between two code generating ends, mutual information measures the number of bits shared between ends. If one end is thought to send a message using a given number of bits the other end is said to receive the message. Any discrepancy in the amount of bits can be attributed to sources unrelated to the communication link such as contaminating noise [10]. Yet the definition of mutual information does not specify sending or receiving side, but measures shared bits in a non-directional manner. The final value depends on the probability of observing a particular combination of codewords. For example, if the appearance of codeword A at endpoint 1 frequently occurs when codeword B appears at endpoint 2, then the (A, B) pair might be an important messaging combination. At this point it might be easy to think that low probability codeword pairs are less relevant than those with high probability. However, information theory says that overly high probability pairs are also uninteresting. We might think that highly repeated patterns might produce desensitization which is described by the concavity of entropy measures. The basic idea is to find a set of codeword pairs carrying most of the bits shared between endpoints.

One of the most basic strategies of data compression known as the Shannon code is to assign the number of bits using  $-\log_2(p_i)$ , where  $p_i$  is the probability of the  $i^{th}$  codeword [11]. While this strategy is not always practical, it is very useful for theoretical illustration. That is because if we average over all possible pairs, i.e. we take the sum  $-\sum p_i \log_2(p_i)$ , we will obtain the entropy which also represents the theoretical minimum number of bits that can describe the data. Notice that  $-0 \cdot \log_2(0) = -1 \cdot \log_2(1) = 0$ , thus high ( $p_i \rightarrow 1$ ) and low ( $p_i \rightarrow 0$ ) probabilities both indicate a low number of bits. This property of entropy indicates that codewords carrying large number of bits exhibit probabilities somewhat in between 0 and 1. The group of codewords fulfilling this expectation forms the typical codeword pairs of the data. This typical set of codewords can be determined by choosing codeword in the vicinity of the entropy.

In the case of brain domains, we identify codewords as the finite set of connectivity patterns, named dFDC states, similar to those illustrated in Fig. 1. Assume now these dFDC states influences the appearance of dFDC states in another part of the brain. Possible outcomes of this end to end relationship are all possible pairs formed by endpoint states. The analysis of such outcomes can be done using joint probabilities  $p_i(X,Y)$  (where  $i$  denote a state pair constituted by a state from endpoint X and another one from state Y) and the joint entropy  $H(X,Y) = -\sum p_i(X,Y) \log_2(p_i(X,Y))$ . Similar to the data compression case, the

relevant end to end state pairs constitute a jointly typical set where  $-\log_2(p_i(X,Y))$  values are in the vicinity of the entropy  $H(X,Y)$ . Using the parameter  $\epsilon > 0$  to represent the vicinity, we can characterize the typical state pairs as

$$|-\log_2(p_i(X,Y)) - H(X,Y)| \leq \epsilon.$$

The theoretical considerations up to this point have delineated a way of selecting state pairs (one state at each brain domain endpoint) representing a relevant number of bits. However, it is unlikely that each brain domain utilizes single state codewords to determine a meaningful message. It is plausible that instead, a sequence of domain states is utilized to establish the messaging between endpoints. While the number of states is undetermined, we shall for now parameterize them as  $n$  and represented as  $(X^n, Y^n)$  indicating pairs of  $n$ -tuple states. An extension to the previous equation would suffice to mathematically describe the typical set [11] (eq. 8.38)

$$|-(1/n) \log_2(p_i(X^n, Y^n)) - H(X,Y)| \leq \epsilon.$$

The picture described by this equation indicate that as domain states appears in sequential progression, the probability of some combinations of states will carry more information (in our case in the form of bits) than others. This panorama is depicted in Fig. 2. In order to estimate an optimal length  $n$  we suggest to use the entropy rate  $H(X,Y)/n$ . This value asymptotically approximates the entropy of the data as  $n$  increases. Using the elbow criteria, an optimal length  $n$  will be at the point where higher lengths does considerable improve the entropy rate.

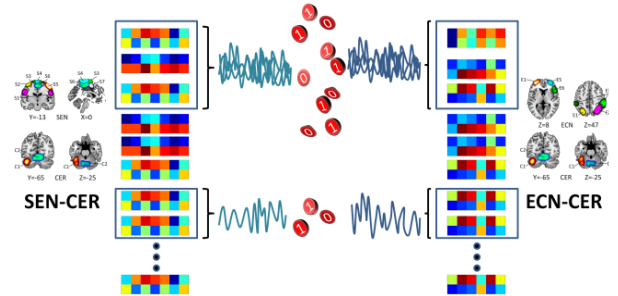


Fig. 2. Concurrent sequences of dFDC states can be associated to different number of bits. The proposed method aims at finding those dFDC state sequences utilizing the theory of jointly typical sets.

## III. MATERIALS AND METHODS

The subject pool consists of 121 healthy subjects (71 females) with a mean age of  $25.4 \pm 8.3$ . A two-sample t-test ( $p > 0.37$ ) show no difference in age between males and females. Subjects did not exhibit injury to the brain, brain-related medical problems, bipolar or psychotic disorders, ADHD or a history of substance abuse/dependence including alcohol. All participants provided informed consent in accordance with institutional guidelines at the University of New Mexico.

Resting state fMRI data was collected from all participants. All images were collected on a 3 Tesla Siemens Trio scanner.

A five-minute resting state run was completed by each participant using a single-shot, gradient-echo echo planar imaging sequence [TR = 2000 ms; TE = 29 ms; flip angle = 75°; FOV = 240 mm; matrix size = 64 x 64]. Foam padding and paper tape was used to restrict motion within the scanner. Thirty-three contiguous, axial 4.55 mm thick slices were selected to provide whole-brain coverage (voxel size: 3.75 x 3.75 x 4.55 mm). The first five images were eliminated to account for T1 equilibrium effects leaving a total of 145 images. Data were preprocessed using the statistical parametric mapping (<http://www.fil.ion.ucl.ac.uk/spm>) software [12]. The preprocessing steps included slice-timing correction, realignment, co-registration, spatial normalization and transformation to the Montreal Neurological Institute (MNI) standard space.

Data were analyzed using Infomax-based group independent component analysis (gICA) [13] with 120 and 100 components for the first and second decomposition levels respectively. A total of 32 components were selected based on frequency content and visual inspection in order to include components that were low noise and free of major artifacts [14]. The gICA time courses were then filtered using a band-pass filter from 0.01 to 0.15 Hz. The set of RSNs were identified and grouped in their functional domains. Spatial overlap with functional brain areas were confirmed by visual comparison with the 90 spatial maps defined by Shirer [15] and by running peak activation coordinates through the meta-analysis software publicly available at <http://www.neurosynth.org/>. The functional domains are SBC (Subcortical), CER (Cerebellar), SEN (Sensorimotor), VIS (Visual), DMN (Default Mode Network) and ECN (Executive Control Network). Chosen RSNs have peak activations in grey matter, low spatial overlap with known vascular, ventricular, motion, and susceptibility artifacts.

#### IV. RESULTS

##### A. Mutual Information

As briefly stated in the introduction, our illustrative goal is to study the code used between cerebrum and cerebellum. For this purpose, the communication with the cerebellum (CER domain) was paired with all other defined domains from the cerebrum defining six domain pairs (ECN-CER, SEN-CER, SBC-CER, CER-CER, VIS-CER and DMN-CER). On each of the 6 dFDCs, we used k-means clustering [16] with a correlation distance metric and a window size of 80 seconds. The number of clusters for each dFDC was set to three after using the elbow criterion on the cluster validity index [7].

The first step is to determine significant mutual information. Mutual information analysis will allow us to determine the domain pairs of interest. There were 15 cross domain pairs for which joint probabilities and mutual information were estimated. Cross-domain mutual information was analyzed utilizing the framework previously proposed in [5]. A null model bootstrapped with ten million iterations was estimated to identify significant mutual information. The results are displayed in Fig. 3. Significant mutual information was found between SEN-CER and ECN-CER dFDCs which indicate a high probability of finding a communication code between these two domains.

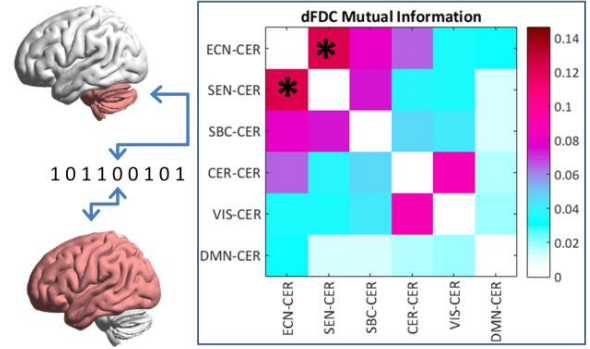


Fig. 3. Mutual information among dFDCs. Only one out of the 15 possible combinations was significant (false discovery rate corrected).

##### B. Results for Jointly Typical Sets

The two endpoints that we utilize are the ECN-CER and SEN-CER. The alphabets for these two endpoints correspond to the estimated domain states that can be seen in Fig. 1 and Fig. 4. Fig. 1 was used both as illustrative example and to show results from this section. Fig. 4 completes the information needed to discuss the significant results found. Applying the elbow criteria to entropy rates, illustrated in Fig. 5a, we estimated an optimal length of  $n = 3$  words.

The next parameter to set is  $\epsilon$ . First, we estimated joint probabilities and joint entropies calculating the joint occurrence of triplets (three temporally consequent domain states) of the ECN-CER and SEN-CER. The procedure of selecting triplets from each domain is depicted in Fig. 2. Fig. 5b displays the difference  $|(1/n) \log_2(p_i(X^n, Y^n)) - H(X, Y)|$  for all triplet combinations found. Notice that not all combinations occurred and thus their estimated probability was zero. These zero probabilities were not in the plot. Instead of setting  $\epsilon$  we choose the three triplet combinations with the lowest value in Fig. 5b. These three sets of triplets hold most of the bits in the whole sequence of dynamic states.

Fig. 6 shows the triplet sequences at each endpoint ECN-CER and SEN-CER. The results indicate that information is higher when ECN-CER is in its state 1 and SEN-CER switch between corresponding state 2 and state 3. Also, state 3 of SEN-CER with a transition state 2 to state 1 in ECN-CER have equal amount of information.

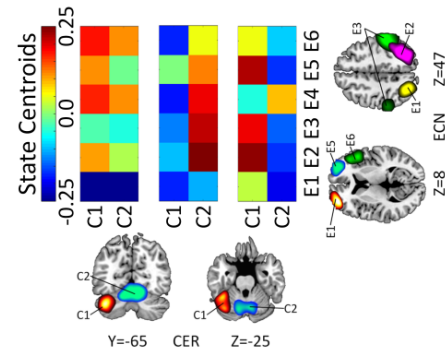


Fig. 4. This dFDC showed significant mutual information with the dFDC displayed in Fig. 1.

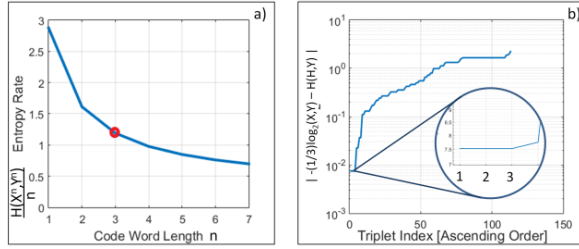


Fig. 5. 5a) Elbow criteria to determine the codeword length  $n$ . 5b) Three different sets of triplets (see Fig. 6) shared the same closeness with the entropy  $H(X,Y)$ .

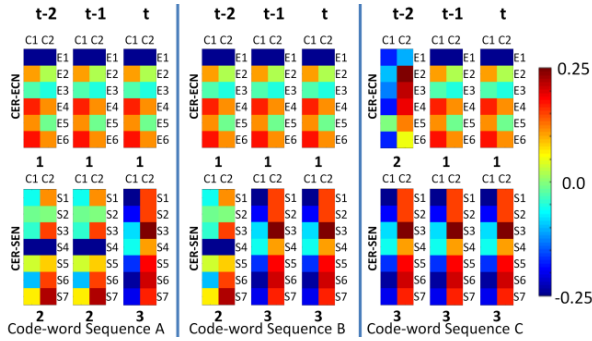


Fig. 6. The three sequence of triplets referred in Fig. 5. These sequences carry most of the bits among ECN, SEN and CER domains.

## V. DISCUSSION

As shown in this work, and demonstrated in a previous publication from our group [5], there is significant information sharing between ECN, cerebellum and sensorimotor brain domains. The novelty in the current analysis is the discovery of those state sequences that characterizes this information. These sequences can be thought as codewords in the communication among these domains. The results show that high information transfer occurs when one of the domain states stay stable while the other one suffer a state transition. This suggests that one domain pair might be executing a function that requires a specific connectivity state. At the same time, the other domain connectivity might have ended its requirement for a given domain state and it is switching to optimally use the time. This connectivity exchange carries most of the bits in the communication and might be important for the healthy functioning of resting state brain.

Although the data was obtained from resting state brains, we can see that information is exchanged among areas that are usually active during demanding tasks such as the ECN and sensorimotor areas [17]. The engagement of these areas is not inconsistent with the resting state brain. These areas might provide an extrospective state of mind within the resting state to provide readiness for external stimuli response and switch out of resting state [18]. Following these considerations, it can be considered important for the resting state brain that significant amounts of information be shared among areas required for attention and task execution.

## ACKNOWLEDGMENT

This work was supported by National Institutes of Health grants 2R01EB005846, R01EB020407 and P20GM103472 and National Science Foundation grants NSF grants 1539067/1631819 to V.D.C.

## REFERENCES

- [1] M. P. van den Heuvel, R. C. Mandl, R. S. Kahn, H. Pol, and E. Hilleke, "Functionally linked resting - state networks reflect the underlying structural connectivity architecture of the human brain," *Human brain mapping*, vol. 30, pp. 3127-3141, 2009.
- [2] D. S. Bassett and E. Bullmore, "Small-world brain networks," *Neuroscientist*, vol. 12, pp. 512-23, Dec 2006.
- [3] M. E. Raichle, "The brain's default mode network," *Annual review of neuroscience*, vol. 38, pp. 433-447, 2015.
- [4] R. Ptak, "The frontoparietal attention network of the human brain action, saliency, and a priority map of the environment," *The Neuroscientist*, vol. 18, pp. 502-515, 2012.
- [5] V. M. Vergara, R. Miller, and V. Calhoun, "An information theory framework for dynamic functional domain connectivity," *J Neurosci Methods*, vol. 284, pp. 103-111, Jun 01 2017.
- [6] A. Abrol, C. Chaze, E. Damaraju, and V. D. Calhoun, "The chronnectome: Evaluating replicability of dynamic connectivity patterns in 7500 resting fMRI datasets," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2016, pp. 5571-5574, Aug 2016.
- [7] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cereb Cortex*, vol. 24, pp. 663-76, Mar 2014.
- [8] M. Yaesoubi, E. A. Allen, R. L. Miller, and V. D. Calhoun, "Dynamic coherence analysis of resting fMRI data to jointly capture state-based phase, frequency, and time-domain information," *Neuroimage*, vol. 120, pp. 133-42, Oct 15 2015.
- [9] R. L. Miller, V. M. Vergara, D. B. Keator, and V. D. Calhoun, "A Method for Intertemporal Functional-Domain Connectivity Analysis: Application to Schizophrenia Reveals Distorted Directional Information Flow," *IEEE Trans Biomed Eng*, vol. 63, pp. 2525-2539, Dec 2016.
- [10] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, p. 3, 2001.
- [11] T. M. Cover and J. A. Thomas, *Elements of information theory*: John Wiley & Sons, 2012.
- [12] K. J. Friston, "Statistical parametric mapping," in *Neuroscience Databases*, ed: Springer, 2003, pp. 237-250.
- [13] V. D. Calhoun and T. Adali, "Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Rev Biomed Eng*, vol. 5, pp. 60-73, 2012.
- [14] E. A. Allen, E. B. Erhardt, E. Damaraju, W. Gruner, J. M. Segall, R. F. Silva, et al., "A baseline for the multivariate comparison of resting-state networks," *Front Syst Neurosci*, vol. 5, p. 2, 2011.
- [15] W. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. Greicius, "Decoding subject-driven cognitive states with whole-brain connectivity patterns," *Cerebral cortex*, vol. 22, pp. 158-165, 2012.
- [16] S. Lloyd, "Least squares quantization in PCM," *IEEE transactions on information theory*, vol. 28, pp. 129-137, 1982.
- [17] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 9673-9678, 2005.
- [18] P. Fransson, "Spontaneous low - frequency BOLD signal fluctuations: An fMRI investigation of the resting - state default mode of brain function hypothesis," *Human brain mapping*, vol. 26, pp. 15-29, 2005.